

19

## SEQUENCE LISTING

## (1) GENERAL INFORMATION:

- (i) APPLICANT: Hinkkanen, Ari
- (ii) TITLE OF INVENTION: A New Fusion Protein and Its Use in an Immunoassay for the Simultaneous Detection of Autoantibodies Related to Insulin-Dependent Diabetes Mellitus
- (iii) NUMBER OF SEQUENCES: 13
- (iv) CORRESPONDENCE ADDRESS:
  - (A) ADDRESSEE: Rothwell, Figg, Ernst & Kurz, P.C.
  - (B) STREET: 555 Thirteenth Street N.W., Suite 701-E
  - (C) CITY: Washington
  - (D) STATE: D.C.
  - (E) COUNTRY: USA
  - (F) ZIP: 20004
- (v) COMPUTER READABLE FORM:
  - (A) MEDIUM TYPE: Floppy disk
  - (B) COMPUTER: IBM PC compatible
  - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
  - (D) SOFTWARE: PatentIn Release #1.0, Version #1.30
- (vi) CURRENT APPLICATION DATA:
  - (A) APPLICATION NUMBER: US 09/015,399
  - (B) FILING DATE: 29-JAN-1998
  - (C) CLASSIFICATION:
- (vii) ATTORNEY/AGENT INFORMATION:
  - (A) NAME: Ihnen, Jeffrey L.
  - (B) REGISTRATION NUMBER: 28,957
  - (C) REFERENCE/DOCKET NUMBER: 2328-111
- (ix) TELECOMMUNICATION INFORMATION:
  - (A) TELEPHONE: 202-783-6040
  - (B) TELEFAX: 202-783-6031

## (2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 8 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS:
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: peptide
- (v) FRAGMENT TYPE: N-terminal
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:  
Asp Tyr Lys Asp Asp Asp Asp Lys  
1 5

## (2) INFORMATION FOR SEQ ID NO:2:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 9 amino acids

20

(B) TYPE: amino acid  
(C) STRANDEDNESS:  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Lys Lys Lys Arg Pro Arg Lys Lys Lys  
1 5

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 10 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS:  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: C-terminal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

Cys Asn Gly Ser His His His His His  
1 5 10

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 9 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS:  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Lys Lys Lys Arg Ser Arg Lys Lys Lys  
1 5

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 979 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS:  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Met Arg Arg Pro Arg Arg Pro Gly Gly Leu Gly Gly Ser Gly Gly Leu  
1 5 10 15

21

Arg Leu Leu Leu Cys Leu Leu Leu Leu Ser Ser Arg Pro Gly Gly Cys  
 20 25 30  
 Ser Ala Val Ser Ala His Gly Cys Leu Phe Asp Arg Arg Leu Cys Ser  
 35 40 45  
 His Leu Glu Val Cys Ile Gln Asp Gly Leu Phe Gly Gln Cys Gln Val  
 50 55 60  
 Gly Val Gly Gln Ala Arg Pro Leu Leu Gln Val Thr Ser Pro Val Leu  
 65 70 75 80  
 Gln Arg Leu Gln Gly Val Leu Arg Gln Leu Met Ser Gln Gly Leu Ser  
 85 90 95  
 Trp His Asp Asp Leu Thr Gln Tyr Val Ile Ser Gln Glu Met Glu Arg  
 100 105 110  
 Ile Pro Arg Leu Arg Pro Pro Glu Pro Arg Pro Arg Asp Arg Ser Gly  
 115 120 125  
 Leu Ala Pro Lys Arg Pro Gly Pro Ala Gly Glu Leu Leu Leu Gln Asp  
 130 135 140  
 Ile Pro Thr Gly Ser Ala Pro Ala Ala Gln His Arg Leu Pro Gln Pro  
 145 150 155 160  
 Pro Val Gly Lys Gly Gly Ala Gly Ala Ser Ser Ser Leu Ser Pro Leu  
 165 170 175  
 Gln Ala Glu Leu Leu Pro Pro Leu Leu Glu His Leu Leu Leu Pro Pro  
 180 185 190  
 Gln Pro Pro His Pro Ser Leu Ser Tyr Glu Pro Ala Leu Leu Gln Pro  
 195 200 205  
 Tyr Leu Phe His Gln Phe Gly Ser Arg Asp Gly Ser Arg Val Ser Glu  
 210 215 220  
 Gly Ser Pro Gly Met Val Ser Val Gly Pro Leu Pro Lys Ala Glu Ala  
 225 230 235 240  
 Pro Ala Leu Phe Ser Arg Thr Ala Ser Lys Gly Ile Phe Gly Asp His  
 245 250 255  
 Pro Gly His Ser Tyr Gly Asp Leu Pro Gly Pro Ser Pro Ala Gln Leu  
 260 265 270  
 Phe Gln Asp Ser Gly Leu Leu Tyr Leu Ala Gln Glu Leu Pro Ala Pro  
 275 280 285  
 Ser Arg Ala Arg Val Pro Arg Leu Pro Glu Gln Gly Ser Ser Ser Arg  
 290 295 300  
 Ala Glu Asp Ser Pro Glu Gly Tyr Glu Lys Glu Gly Leu Gly Asp Arg  
 305 310 315 320  
 Gly Glu Lys Pro Ala Ser Pro Ala Val Gln Pro Asp Ala Ala Leu Gln  
 325 330 335  
 Arg Leu Ala Ala Val Leu Ala Gly Tyr Gly Val Glu Leu Arg Gln Leu  
 340 345 350  
 Thr Pro Glu Gln Leu Ser Thr Leu Leu Thr Leu Leu Gln Leu Leu Pro  
 355 360 365

22

Lys Gly Ala Gly Arg Asn Pro Gly Gly Val Val Asn Val Gly Ala Asp  
370 375 380

Ile Lys Lys Thr Met Glu Gly Pro Val Glu Gly Arg Asp Thr Ala Glu  
385 390 395 400

Leu Pro Ala Arg Thr Ser Pro Met Pro Gly His Pro Thr Ala Ser Pro  
405 410 415

Thr Ser Ser Glu Val Gln Gln Val Pro Ser Pro Val Ser Ser Glu Pro  
420 425 430

Pro Lys Ala Ala Arg Pro Pro Val Thr Pro Val Leu Leu Glu Lys Lys  
435 440 445

Ser Pro Leu Gly Gln Ser Gln Pro Thr Val Ala Gly Gln Pro Ser Ala  
450 455 460

Arg Pro Ala Ala Glu Glu Tyr Gly Tyr Ile Val Thr Asp Gln Lys Pro  
465 470 475 480

Leu Ser Leu Ala Ala Gly Val Lys Leu Leu Glu Ile Leu Ala Glu His  
485 490 495

Val His Met Ser Ser Gly Ser Phe Ile Asn Ile Ser Val Val Gly Pro  
500 505 510

Ala Leu Thr Phe Arg Ile Arg His Asn Glu Gln Asn Leu Ser Leu Ala  
515 520 525

Asp Val Thr Gln Gln Ala Gly Leu Val Lys Ser Glu Leu Glu Ala Gln  
530 535 540

Thr Gly Leu Gln Ile Leu Gln Thr Gly Val Gly Gln Arg Glu Glu Ala  
545 550 555 560

Ala Ala Val Leu Pro Gln Thr Ala His Ser Thr Ser Pro Met Arg Ser  
565 570 575

Val Leu Leu Thr Leu Val Ala Leu Ala Gly Val Ala Gly Leu Leu Val  
580 585 590

Ala Leu Ala Val Ala Leu Cys Val Arg Gln His Ala Arg Gln Gln Asp  
595 600 605

Lys Glu Arg Leu Ala Ala Leu Gly Pro Glu Gly Ala His Gly Asp Thr  
610 615 620

Thr Phe Glu Tyr Gln Asp Leu Cys Arg Gln His Met Ala Thr Lys Ser  
625 630 635 640

Leu Phe Asn Arg Ala Glu Gly Pro Pro Glu Pro Ser Arg Val Ser Ser  
645 650 655

Val Ser Ser Gln Phe Ser Asp Ala Ala Gln Ala Ser Pro Ser Ser His  
660 665 670

Ser Ser Thr Pro Ser Trp Cys Glu Glu Pro Ala Gln Ala Asn Met Asp  
675 680 685

Ile Ser Thr Gly His Met Ile Leu Ala Tyr Met Glu Asp His Leu Arg  
690 695 700

Asn Arg Asp Arg Leu Ala Lys Glu Trp Gln Ala Leu Cys Ala Tyr Gln  
705 710 715 720

23

Ala Glu Pro Asn Thr Cys Ala Thr Ala Gln Gly Glu Gly Asn Ile Lys  
725 730 735

Lys Asn Arg His Pro Asp Phe Leu Pro Tyr Asp His Ala Arg Ile Lys  
740 745 750

Leu Lys Val Glu Ser Ser Pro Ser Arg Ser Asp Tyr Ile Asn Ala Ser  
755 760 765

Pro Ile Ile Glu His Asp Pro Arg Met Pro Ala Tyr Ile Ala Thr Gln  
770 775 780

Gly Pro Leu Ser His Thr Ile Ala Asp Phe Trp Gln Met Val Trp Glu  
785 790 795 800

Ser Gly Cys Thr Val Ile Val Met Leu Thr Pro Leu Val Glu Asp Gly  
805 810 815

Val Lys Gln Cys Asp Arg Tyr Trp Pro Asp Glu Gly Ala Ser Leu Tyr  
820 825 830

His Val Tyr Glu Val Asn Leu Val Ser Glu His Ile Trp Cys Glu Asp  
835 840 845

Phe Leu Val Arg Ser Phe Tyr Leu Lys Asn Val Gln Thr Gln Glu Thr  
850 855 860

Arg Thr Leu Thr Gln Phe His Phe Leu Ser Trp Pro Ala Glu Gly Thr  
865 870 875 880

Pro Ala Ser Thr Arg Pro Leu Leu Asp Phe Arg Arg Lys Val Asn Lys  
885 890 895

Cys Tyr Arg Gly Arg Ser Cys Pro Ile Ile Val His Cys Ser Asp Gly  
900 905 910

Ala Gly Arg Thr Gly Thr Tyr Ile Leu Ile Asp Met Val Leu Asn Arg  
915 920 925

Met Ala Lys Gly Val Lys Glu Ile Asp Ile Ala Ala Thr Leu Glu His  
930 935 940

Val Arg Asp Gln Arg Pro Gly Leu Val Arg Ser Lys Asp Gln Phe Glu  
945 950 955 960

Phe Ala Leu Thr Ala Val Ala Glu Glu Val Asn Ala Ile Leu Lys Ala  
965 970 975

Leu Pro Gln

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 585 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS:  
(D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Met Ala Ser Pro Gly Ser Gly Phe Trp Ser Phe Gly Ser Glu Asp Gly  
1 5 10 15

24

Ser Gly Asp Ser Glu Asn Pro Gly Thr Ala Arg Ala Trp Cys Gln Val  
 20 25 30  
 Ala Gln Lys Phe Thr Gly Gly Ile Gly Asn Lys Leu Cys Ala Leu Leu  
 35 40 45  
 Tyr Gly Asp Ala Glu Lys Pro Ala Glu Ser Gly Gly Ser Gln Pro Pro  
 50 55 60  
 Arg Ala Ala Ala Arg Lys Ala Ala Cys Ala Cys Asp Gln Lys Pro Cys  
 65 70 75 80  
 Ser Cys Ser Lys Val Asp Val Asn Tyr Ala Phe Leu His Ala Thr Asp  
 85 90 95  
 Leu Leu Pro Ala Cys Asp Gly Glu Arg Pro Thr Leu Ala Phe Leu Gln  
 100 105 110  
 Asp Val Met Asn Ile Leu Leu Gln Tyr Val Val Lys Ser Phe Asp Arg  
 115 120 125  
 Ser Thr Lys Val Ile Asp Phe His Tyr Pro Asn Glu Leu Leu Gln Glu  
 130 135 140  
 Tyr Asn Trp Glu Leu Ala Asp Gln Pro Gln Asn Leu Glu Glu Ile Leu  
 145 150 155 160  
 Met His Cys Gln Thr Leu Lys Tyr Ala Ile Lys Thr Gly His Pro  
 165 170 175  
 Arg Tyr Phe Asn Gln Leu Ser Thr Gly Leu Asp Met Val Gly Leu Ala  
 180 185 190  
 Ala Asp Trp Leu Thr Ser Thr Ala Asn Thr Asn Met Phe Thr Tyr Glu  
 195 200 205  
 Ile Ala Pro Val Phe Val Leu Leu Glu Tyr Val Thr Leu Lys Lys Met  
 210 215 220  
 Arg Glu Ile Ile Gly Trp Pro Gly Gly Ser Gly Asp Gly Ile Phe Ser  
 225 230 235 240  
 Pro Gly Gly Ala Ile Ser Asn Met Tyr Ala Met Met Ile Ala Arg Phe  
 245 250 255  
 Lys Met Phe Pro Glu Val Lys Glu Lys Gly Met Ala Ala Leu Pro Arg  
 260 265 270  
 Leu Ile Ala Phe Thr Ser Glu His Ser His Phe Ser Leu Lys Lys Gly  
 275 280 285  
 Ala Ala Ala Leu Gly Ile Gly Thr Asp Ser Val Ile Leu Ile Lys Cys  
 290 295 300  
 Asp Glu Arg Gly Lys Met Ile Pro Ser Asp Leu Glu Arg Arg Ile Leu  
 305 310 315 320  
 Glu Ala Lys Gln Lys Gly Phe Val Pro Phe Leu Val Ser Ala Thr Ala  
 325 330 335  
 Gly Thr Thr Val Tyr Gly Ala Phe Asp Pro Leu Leu Ala Val Ala Asp  
 340 345 350  
 Ile Cys Lys Lys Tyr Lys Ile Trp Met His Val Asp Ala Ala Trp Gly  
 355 360 365

25

Gly Gly Leu Leu Met Ser Arg Lys His Lys Trp Lys Leu Ser Gly Val  
 370 375 380  
 Glu Arg Ala Asn Ser Val Thr Trp Asn Pro His Lys Met Met Gly Val  
 385 390 395 400  
 Pro Leu Gln Cys Ser Ala Leu Leu Val Arg Glu Glu Gly Leu Met Gln  
 405 410 415  
 Asn Cys Asn Gln Met His Ala Ser Tyr Leu Phe Gln Gln Asp Lys His  
 420 425 430  
 Tyr Asp Leu Ser Tyr Asp Thr Gly Asp Lys Ala Leu Gln Cys Gly Arg  
 435 440 445  
 His Val Asp Val Phe Lys Leu Trp Leu Met Trp Arg Ala Lys Gly Thr  
 450 455 460  
 Thr Gly Phe Glu Ala His Val Asp Lys Cys Leu Glu Leu Ala Glu Tyr  
 465 470 475 480  
 Leu Tyr Asn Ile Ile Lys Asn Arg Glu Gly Tyr Glu Met Val Phe Asp  
 485 490 495  
 Gly Lys Pro Gln His Thr Asn Val Cys Phe Trp Tyr Ile Pro Pro Ser  
 500 505 510  
 Leu Arg Thr Leu Glu Asp Asn Glu Glu Arg Met Ser Arg Leu Ser Lys  
 515 520 525  
 Val Ala Pro Val Ile Lys Ala Arg Met Met Glu Tyr Gly Thr Thr Met  
 530 535 540  
 Val Ser Tyr Gln Pro Leu Gly Asp Lys Val Asn Phe Phe Arg Met Val  
 545 550 555 560  
 Ile Ser Asn Pro Ala Ala Thr His Gln Asp Ile Asp Phe Leu Ile Glu  
 565 570 575  
 Glu Ile Glu Arg Leu Gly Gln Asp Leu  
 580 585

## (2) INFORMATION FOR SEQ ID NO:7:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 110 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS:
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

Met Ala Leu Trp Met Arg Leu Leu Pro Leu Leu Ala Leu Leu Ala Leu  
 1 5 10 15  
 Trp Gly Pro Asp Pro Ala Ala Ala Phe Val Asn Gln His Leu Cys Gly  
 20 25 30  
 Ser His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly Phe  
 35 40 45

26

Phe Tyr Thr Pro Lys Thr Arg Arg Glu Ala Glu Asp Leu Gln Val Gly  
 50 55 60

Gln Val Glu Leu Gly Gly Gly Pro Gly Ala Gly Ser Leu Gln Pro Leu  
 65 70 75 80

Ala Leu Glu Gly Ser Leu Gln Lys Arg Gly Ile Val Glu Gln Cys Cys  
 85 90 95

Thr Ser Ile Cys Ser Leu Tyr Gln Leu Glu Asn Tyr Cys Asn  
 100 105 110

## (2) INFORMATION FOR SEQ ID NO:8:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2457 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: cDNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

ACCCGCCCTC GCCGCTCGGC CCCGCGCGTC CCCGCGCGTG CCCTCCTCCC GCCACACGGC 60

ACGCACGCGC GCGCAGGGCC AAGCCGAGGC AGCCGCCCGC AGCTCGCACT CGCTGGCGAC 120

CTGCTCCAGT CTCCAAAGCC GATGGCATCT CCGGGCTCTG GCTTTTGGTC TTTCGGGTCTG 180

GAAGATGGCT CTGGGGATTC CGAGAATCCC GGCACAGCGC GAGCCTGGTG CCAAGTGGCT 240

CAGAAGTTCA CGGGCGGCAT CGGAAACAAA CTGTGCGCCC TGCTCTACGG AGACGCCGAG 300

AAGCCGGCGG AGAGCGGCGG GAGCCAACCC CCGCGGGCCG CCGCCCGGAA GGCCGCCTGC 360

GCTGCGACC AGAAGCCCTG CAGCTGCTCC AAAGTGGATG TCAACTACGC GTTTCTCCAT 420

GCAACAGACC TGCTGCCGGC GTGTGATGGA GAAAGGCCCA CTTTGGCGTT TCTGCAAGAT 480

GTTATGAACA TTTTACTTCA GTATGTGGTG AAAAGTTTCG ATAGATCAAC CAAAGTGATT 540

GATTTCCATT ATCCTAATGA GCTTCTCCAA GAATATAATT GGGAAATTGGC AGACCAACCA 600

CAAAATTTGG AGGAAATTTT GATGCATTGC CAAACAACCTC TAAAATATGC AATTAAAACA 660

GGGCATCCTA GATACTTCAA TCAACTTTCT ACTGGTTTGG ATATGGTTGG ATTAGCAGCA 720

GACTGGCTGA CATCAACAGC AAATACTAAC ATGTTACCTT ATGAAATTGC TCCAGTATTT 780

GTGCTTTTGG AATATGTCAC ACTAAAGAAA ATGAGAGAAA TCATTGGCTG GCCAGGGGGC 840

TCTGGCGATG GGATATTTTC TCCCGGTGGC GCCATATCTA ACATGTATGC CATGATGATC 900

GCACGCTTTA AGATGTTCCC AGAAGTCAAG GAGAAAGGAA TGGCTGCTCT TCCCAGGCTC 960

ATTGCCTTCA CGTCTGAACA TAGTCATTTT TCTCTCAAGA AGGGAGCTGC AGCCTTAGGG 1020

ATTGGAACAG ACAGCGTGAT TCTGATTAAA TGTGATGAGA GAGGGAAAAT GATTCCATCT 1080

GATCTTGAAG GAAGGATTCT TGAAGCCAAA CAGAAAGGGT TTGTTCTTTT CCTCGTGAGT 1140

GCCACAGCTG GAACCACCGT GTACGGAGCA TTTGACCCCC TCTTAGCTGT CGCTGACATT 1200



27

TGCAAAAAGT ATAAGATCTG GATGCATGTG GATGCAGCTT GGGGTGGGGG ATTACTGATG 1260  
 TCCCGAAAAC ACAAGTGGAA ACTGAGTGGC GTGGAGAGGG CCAACTCTGT GACGTGGAAT 1320  
 CCACACAAGA TGATGGGAGT CCCTTTGCAG TGCTCTGCTC TCCTGGTTAG AGAAGAGGGA 1380  
 TTGATGCAGA ATTGCAACCA AATGCATGCC TCCTACCTCT TTCAGCAAGA TAAACATTAT 1440  
 GACCTGTCCT ATGACACTGG AGACAAGGCC TTACAGTGCG GACGCCACGT TGATGTTTTT 1500  
 AAACTATGGC TGATGTGGAG GGCAGAGGGG ACTACCGGGT TTGAAGCGCA TGTGATAAA 1560  
 TGTTTGGAGT TGGCAGAGTA TTTATACAAC ATCATAAAAA ACCGAGAAGG ATATGAGATG 1620  
 GTGTTTGATG GGAAGCCTCA GCACACAAAT GTCTGCTTCT GGTACATTCC TCCAAGCTTG 1680  
 CGTACTCTGG AAGACAATGA AGAGAGAATG AGTCGCCTCT CGAAGGTGGC TCCAGTGATT 1740  
 AAAGCCAGAA TGATGGAGTA TGAACCACA ATGGTCAGCT ACCAACCCTT GGGAGACAAG 1800  
 GTCAATTTCT TCCGCATGGT CATCTCAAAC CCAGCGGCAA CTCACCAAGA CATTGACTTC 1860  
 CTGATTGAAC AAATAGAACG CCTTGGACAA GATTATAAT AAQCTTGCTC ACCAAGCTGT 1920  
 TCCACTTCTC TAGAGAACAT GCCCTCAGCT AAGCCCCCTA CTGAGAAACT TCCTTTGAGA 1980  
 ATTGTGCGAC TTCACAAAAT GCAAGGTGAA CACCACTTTG TCTCTGAGAA CAGACGTTAC 2040  
 CAATTATGGA GTGTCAACCAG CTGCCAAAAT CGTAGGTGTT GGCTCTGCTG GTCAGTGGAG 2100  
 TAGTTGCTAC TCTTCAGAAT ATGGACAAAG AAGGCACAGG TGTAATATA GTAGCAGGAT 2160  
 GAGGAACCTC AAACTGGGTA TCATTTGCAC GTGCTCTTCT GTTCTCAAAT GCTAAATGCA 2220  
 AACACTGTGT ATTTATTAGT TAGGTGTGCC AAACTACCGT TCCCAAATTG GTGTTTCTGA 2280  
 ATGACATCAA CATTCCCCCA ACATTACTCC ATTACTAAAG ACAGAAAAAA ATAAAAACAT 2340  
 AAAATATACA AACATGTGGC AACCTGTTCT TCCTACCAA TATAAACTTG TGTATGATCC 2400  
 AAGTATTTTA TCTGTGTTGT CTCTCTAAAC CCAAATAAAT GTGTAAATGT GGACACA 2457

## (2) INFORMATION FOR SEQ ID NO:9:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 3613 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

CAGCCCTCT GGCAGGCTCC CGCCAGCGTC GCTGCGGCTC CGGCCCGGGA GCGAGCGCCC 60  
 GGAGCTCGGA AAGATGCGGC GCGCGGCG GCCTGGGGGT CTCGGGGGAT CCGGGGGTCT 120  
 CCGGCTGCTC CTCTGCCTCC TGCTGCTGAG CAGCCGCCCG GGGGGCTGCA GCGCCGTTAG 180  
 TGCCACGGC TGTCTATTG ACCGAGGCT CTGCTCTCAC CTGGAAGTCT GTATTCAGGA 240  
 TGGCTTGTTT GGGCAGTGCC AGGTGGGAGT GGGGCAGGCC CGGCCCTTT TGCAAGTCAC 300

28

CTCCCCAGTT CTCCAACGCT TACAAGGTGT GCTCCGACAA CTCATGTCCC AAGGATTGTC	360
CTGGCACGAT GACCTCACCC AGTATGTGAT CTCTCAGGAG ATGGAGCGCA TCCCCAGGCT	420
TCGCCCCCCA GAGCCCCGTC CAAGGGACAG GTCTGGCTTG GCACCCAAGA GACCTGGTCC	480
TGCTGGAGAG CTGCTTTTAC AGGACATCCC CACTGGCTCC GCCCCTGCTG CCCAGCATCG	540
GCTTCCACAA CCACCAGTGG GCAAAGGTGG AGCTGGGGCC AGCTCCTCTC TGTCCCCTCT	600
GCAGGCTGAG CTGCTCCCGC CTCTCTTGGA GCACCTGCTG CTGCCCCAC AGCCTCCCCA	660
CCCTTCACTG AGTTACGAAC CTGCCTTGCT GCAGCCCTAC CTGTTCCACC AGTTTGCTC	720
CCGTGATGGC TCCAGGGTCT CAGAGGGCTC CCCAGGGATG GTCAGTGTG GCCCCCTGCC	780
CAAGGCTGAA GCCCCTGCCC TCTTCAGCAG AACTGCCTCC AAGGGCATAT TTGGGGACCA	840
CCCTGGCCAC TCCTACGGGG ACCTTCCAGG GCCTTCACCT GCCCAGCTTT TTCAAGACTC	900
TGGGCTGCTC TATCTGGCCC AGGAGTTGCC AGCACCCAGC AGGGCCAGGG TGCCAAGGCT	960
GCCAGAGCAA GGGAGCAGCA GCCGGGCAGA GGACTCCCCA GAGGGCTATG AGAAGGAAGG	1020
ACTAGGGGAT CGTGGAGAGA AGCCTGCTTC CCCAGCTGTG CAGCCAGATG CGGCTCTGCA	1080
GAGGCTGGCC GCTGTGCTGG CGGGCTATGG GGTAGAGCTG CGTCAGCTGA CCCCTGAGCA	1140
GCTCTCCACA CTCCTGACCC TGCTGCAGCT ACTGCCAAG GGTGCAGGAA GAAATCCGGG	1200
AGGGGTTGTA AATGTTGGAG CTGATATCAA GAAAACAATG GAGGGGCCGG TGGAGGGCAG	1260
AGACACAGCA GAGCTTCCAG CCCGCACATC CCCCATGCCT GGACACCCCA CTGCCAGCCC	1320
TACCTCCAGT GAAGTCCAGC AGGTGCCAAG CCCTGTCTCC TCTGAGCTC CCAAAGCTGC	1380
CAGACCCCT GTGACACCTG TCCTGCTAGA GAAGAAAAGC CCACTGGGCC AGAGCCAGCC	1440
CACGGTGGCA GGACAGCCCT CAGCCCGCCC AGCAGCAGAG GAATATGGCT ACATCGTCAC	1500
TGATCAGAAG CCCCTGAGCC TGGCTGCAGG AGTGAAGCTG CTGGAGATCC TGGCTGAGCA	1560
TGTGCACATG TCCTCAGGCA GCTTCATCAA CATCAGTGTG GTGGGACCAG CCCTCACCTT	1620
CCGCATCCGG CACAATGAGC AGAACCTGTC TTTGGCTGAT GTGACCCAAC AAGCAGGGCT	1680
GGTGAAGTCT GAACTGGAAG CACAGACAGG GCTCCAAATC TTGCAGACAG GAGTGGGACA	1740
GAGGGAGGAG GCAGCTGCAG TCCTTCCCCA AACTGCGCAC AGCACCTCAC CCATGCGCTC	1800
AGTGCTGCTC ACTCTGGTGG CCCTGGCAGG TGTGGCTGGG CTGCTGGTGG CTCTGGCTGT	1860
GGCTCTGTGT GTGCGGCAGC ATGCGCGGCA GCAAGACAAG GAGCGCTGG CAGCCCTGGG	1920
GCCTGAGGGG GCCCATGGTG AACTACCTT TGAGTACCAG GACCTGTGCC GCCAGCACAT	1980
GGCCACGAAG TCCTTGTTC AACCAGGAGA GGGTCCACCG GAGCCTTCAC GGGTGAGCAG	2040
TGTGCTCTCC CAGTTCAGCG ACGCAGCCCA GGCCAGCCCC AGCTCCCA CA GCAGCACCCC	2100
GTCTGGTGC GAGGAGCCGG CCCAAGCCAA CATGGACATC TCCACGGGAC ACATGATTCT	2160
GGCATACATG GAGGATCACC TCGGAACCG GGACCGCCTT GCCAAGGAGT GGCAGGCCCT	2220
CTGTGCTTAC CAAGCAGAGC CAAACACCTG TGCCACCGCG CAGGGGGAGG GCAACATCAA	2280

29

AAAGAACCGG CATCCTGACT TCCTGCCCTA TGACCATGCC CGCATAAAAC TGAAGGTGGA 2340  
 GAGCAGCCCT TCTCGGAGCG ATTACATCAA CGCCAGCCCC ATTATTGAGC ATGACCCTCG 2400  
 GATGCCAGCC TACATAGCCA CGCAGGGCCC GCTGTCCCAT ACCATCGCAG ACTTCTGGCA 2460  
 GATGGTGTGG GAGAGCGGCT GCACCGTCAT CGTCATGCTG ACCCGGCTGG TGGAGGATGG 2520  
 TGTCAAGCAG TGTGACCGCT ACTGGCCAGA TGAGGGTGCC TCCCTCTACC ACGTATATGA 2580  
 GGTGAACCTG GTGTCGGAGC ACATCTGGTG CGAGGACTTT CTGGTGCGGA GCTTCTACCT 2640  
 GAAGAACGTG CAGACCCAGG AGACGCGCAC GCTCACGCAG TTCCACTTCC TCAGCTGGCC 2700  
 GGCAGAGGGC ACACCGGCCT CCACGCGGCC CCTGCTGGAC TTCCGCAGGA AGGTGAACAA 2760  
 GTGCTACCGG GGCCGCTCCT GCCCCATCAT CGTGCACTGC AGTGATGGTG CGGGGAGGAC 2820  
 CGGCACCTAC ATCCTCATCG ACATGGTCCT GAACCGCATG GCAAAGGAG TGAAGGAGAT 2880  
 TGACATCGCT GCCACCCTGG AGCATGTCCG TGACCAGCGG CCTGGCCTTG TCGGCTCTAA 2940  
 GGACCAGTTT GAATTTGCCC TGACAGCCGT GCGGGAGGAA GTGAATGCCA TCCTCAAGGC 3000  
 CCTGCCCCAG TGAGACCCTG GGGCCCCCTG GCGGGCAGCC CAGCCTCTGT CCCTCTTTGC 3060  
 CTGTGTGAGC ATCTCTGTGT ACCCACTCCT CACTGCCCCA CCAGCCACCT CTTGGGCATG 3120  
 CTCAGCCCTT CCTAGAAGAG TCAGGAAGGG AAAGCCAGAA GGGGCACGCC TGCCAGCCCT 3180  
 CGCATCCCAG AGCCTGGGGC ATCCCAGAGC CCAGGGCATC CCATGGGGGT GCTGCAGCCA 3240  
 GGAGGAGAGG AAAGGACATG GGTAGCAATT CTACCCAGAG CCTTCTCCTG CCTACATTCC 3300  
 CTGGCCTGGC TCTCCTGTAG CTCTCCTGGG GTTCTGGGAG TTCCCTGAAC ATCTGTGTGT 3360  
 GTCCCCCTAT GCTCCAGTAT GGAAGAATGG GGTGGAGGGT CGCCACACCC GGCTCCCCCT 3420  
 GCTTCTCAGC CCCGGGCCTG CCTCTGACTC ACACTTGGGC GCTCTGCCCT CCCTGGCCTC 3480  
 ACGCCAGCC TGGTCCCACC ACCCTCCCAC CATGCGCTGC TCAACCTCTC TCCTTCTGGC 3540  
 GCAAGAGAAC ATTTCTAGAA AAAACTACTT TTGTACCAGT GTGAATAAAG TTAGTGTGTT 3600  
 GTCTGTGCAG CTG 3613

## (2) INFORMATION FOR SEQ ID NO:10:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 4992 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

CTCGAGGGGC CTAGACATTG CCCTCCAGAG AGAGCACCCA ACACCTCCA GGCTTGACCG 60  
 GCAGGGTGT CCCCTTCTA CCTTGAGAG AGCAGCCCCA GGGCATCCTG CAGGGGGTGC 120  
 TGGACACCA GCTGGCCTTC AAGGTCTCTG CCTCCCTCCA GCCACCCAC TACACGCTGC 180

30

TGGGATCCTG GATCTCAGCT CCCTG6CCGA CAACACTGGC AAACCTCCTAC TCATCCACGA	240
AGGCCCTCCT GGGCATGGTG GTCCTTCCCA GCCTGGCAGT CTGTTCCCTCA CACACCTTGT	300
TAGTGCCCAG CCCCTGAGGT TGCAGCTGGG GGTGTCTCTG AAGGGCTGTG AGCCCCCAGG	360
AAGCCCTGGG GAAGTGCTG CCTTGCCCTCC CCCCAGCCCT GCCAGCGCCT GGCTCTGCCC	420
TCCTACCTGG GCTCCCCCA TCCAGCCTCC CTCCCTACAC ACTCCTCTCA AGGAGGCACC	480
CATGTCTCTT CCAGCTGCCG GGCTCAGAG CACTGTGGCG TCCTGGGGCA GCCACCGCAT	540
GTCTGCTGT GGCATGGCTC AGGGTGGAAA GGGCGGAAGG GAGGGGTCTT GCAGATAGCT	600
GGTGCCCACT ACCAAACCCG CTCGGGGCAG GAGAGCCAAA GGCTGGGTGT GTGCAGAGCG	660
GCCCCGAGAG GTTCCGAGGC TGAGGCCAGG GTGGGACATA GGGATGCGAG GGGCCGGGGC	720
ACAGGATACT CCAACCTGCC TCCCCCATG GTCTCATCCT CCTGCTTCTG GGACCTCCTG	780
ATCCTGCCCC TGGTGCTAAG AGGCAGGTAA GGGGCTGCAG GCAGCAGGGC TCGGAGCCCA	840
TGCCCCCTCA CCATGGGTCA GGCTGGACCT CCAGGTGCCT GTTCTGGGGA GCTGGGAGGG	900
CCGGAGGGGT GTACCCAGG GGCTCAGCCC AGATGACACT ATGGGGGTGA TGGTGTCTG	960
GGACCTGGCC AGGAGAGGGG AGATGGGCTC CCAGAAGAGG AGTGGGGGCT GAGAGGGTGC	1020
CTGGGGGGCC AGGACGGAGC TGGGCCAGTG CACAGCTTCC CACACCTGCC CACCCCCAGA	1080
GTCTGCCGC CACCCCCAGA TCACACGGAA GATGAGGTCC GAGTGGCCTG CTGAGGACTT	1140
GCTGCTTGTG CCCAGGTCCC CAGGTCATGC CCTCCTTCTG CCACCTGGG GAGCTGAGGG	1200
CCTCAGCTGG GGCTGCTGTC CTAAGGCAGG GTGGGAACCTA GGCAGCCAGC AGGGAGGGGA	1260
CCCCCTCCTC ACTCCCACTC TCCACCCCC ACCACCTTGG CCCATCCATG GCGGCATCTT	1320
GGGCCATCCG GGAAGTGGGA CAGGGGTCTT GGGGACAGGG GTCCGGGGAC AGGGTCTCTG	1380
GGACAGGGGT GTGGGGACAG GGGTCTGGGG ACAGGGGTGT GGGGACAGGG GTGTGGGGAC	1440
AGGGGTCTTG GGACAGGGGT GTGGGGACAG GGGTCCGGGG ACAGGGGTGT GGGGACAGGG	1500
GTCTGGGGAC AGGGGTGTGG GGACAGGGGT GTGGGGACAG GGGTCTGGGG ACAGGGGTGT	1560
GGGGACAGGG GTCTGGGGGA CAGGGGTCTG GGGACAGGGG TGTGGGGACA GGGGTGTGGG	1620
GACAGGGGTG TGGGGACAGG GGTCTGGGG ATAGGGGTGT GGGGACAGGG GTGTGGGGAC	1680
AGGGGTCCCG GGGACAGGGG TGTGGGGACA GGGGTGTGGG GACAGGGGTC CTGGGGACAG	1740
GGGTCTGAGG ACAGGGGTGT GGGACAGGG GTCTGGGGGA CAGGGGTCTT GGGGACAGGG	1800
GTCTGGGGGA CAGGGGTCTG GGGACAGCAG CGCAAAGAGC CCCGCCCTGC AGCCTCCAGC	1860
TCTCTGGTGC TAATGTGGAA AGTGGCCCAG GTGAGGGCTT TGCTCTCTCTG GAGACATTTG	1920
CCCCAGCTG TGAGCAGGGA CAGGTCTGGC CACCGGGCCC CTGGTTAAGA CTCTAATGAC	1980
CCGCTGGTCC TGAGGAAGAG GTGCTGACGA CCAAGGAGAT CTTCCACAG ACCCAGCACC	2040
AGGGAAATGG TCCGGAATT GCAGCCTCAG CCCCAGCCA TCTGCCGACC CCCCCACCCC	2100
GCCCTAATGG GCCAGGCGGC AGGGGTGAC AGGTAGGGGA GATGGGCTCT GAGACTATAA	2160

31

AGCCAGCGGG GGCCAGCAG CCCTCAGCCC TCCAGGACAG GCTGCATCAG AAGAGGCCAT	2220
CAAGCAGGTC TGTTCCAAGG GCCTTTGCGT CAGGTGGGCT CAGGGTTCCA GGGTGGCTGG	2280
ACCCAGGCC CCAGCTCTGC AGCAGGGAGG ACGTGGCTGG GCTCGTGAAG CATGTGGGGG	2340
TGAGCCCAGG GGCCCCAAGG CAGGGCACCT GGCCTTCAGC CTGCCTCAGC CCTGCCTGTC	2400
TCCCAGATCA CTGTCCTTCT GCCATGGCCC TGTGGATGCG CTTCTGCCC CTGCTGGCGC	2460
TGCTGGCCCT CTGGGGACCT GACCCAGCCG CAGCCTTGT GAACCAACAC CTGTGCGGCT	2520
CACACCTGGT GGAAGCTCTC TACCTAGTGT GCGGGGAACG AGGCTTCTTC TACACACCCA	2580
AGACCCGCCG GGAGGCAGAG GACCTGCAGG GTGAGCCAAC CGCCATTGC TGCCCTGGC	2640
CGCCCCAGC CACCCCTGC TCCTGGCGCT CCCACCCAGC ATGGGCAGAA GGGGCAGGA	2700
GGCTGCCACC CAGCAGGGGG TCAGGTGCAC TTTTAAAGG AGAAGTTCTC TTGGTCACGT	2760
CCTAAAAGTG ACCAGCTCCC TGTGGCCAG TCAGAATCTC AGCCTGAGGA CGGTGTTGGC	2820
FTCGGCAGCC CCGAGATACA TCAGAGGGTG GGCACGCTCC TCCCTCCACT CGCCCTCAA	2880
ACAAATGCCC CGCAGCCCAT TTCTCCACCC TCATTTGATG ACCGCAGATT CAAGTGTTTT	2940
GTTAAGTAAA GTCCTGGGTG ACCTGGGGTC ACAGGGTGCC CCACGCTGCC TGCTCTGGG	3000
CGAACACCCC ATCACGCCG GAGGAGGGCG TGGCTGCCTG CCTGAGTGGG CCAGACCCCT	3060
GTCGCCAGCC TCACGGCAGC TCCATAGTCA GGAGATGGGG AAGATGCTGG GSACAGGCC	3120
TGGGGAGAAG TACTGGGATC ACCTGTTCAG GCTCCCACTG TGACGCTGCC CCGGGGCGGG	3180
GGAAGGAGGT GGGACATGTG GCGTTGGGG CCGTAGGTG CACACCCAGT GTGGGTGACC	3240
CTCCCTCTAA CCTGGGTCCA GCGCGCTGG AGATGGGTGG GAGTGCAGC TAGGGCTGGC	3300
GGGCAGGCGG GCACTGTGTC TCCCTGACTG TGTCTCTCTG TGTCCCTCTG CCTGCGCGCT	3360
GTTCCGGAAC CTGCTCTGCG CGGCACGTCC TGGCAGTGGG GCAGGTGGAG CTGGGCGGGG	3420
GCCCTGGTGC AGGCAGCCTG CAGCCCTTGG CCCTGGAGGG GTCCCTGCAG AAGCGTGGCA	3480
TTGTGGAACA ATGCTGTACC AGCATCTGCT CCCTCTACCA GCTGGAGAAC TACTGCAACT	3540
AGACGCAGCC TGCAGGCAGC CCCACACCCG CCGCCTCCTG CACCGAGAGA GATGGAATAA	3600
AGCCCTTGAA CCAGCCCTGC TGTGCCGTCT GTGTGTCTTG GGGGCCCTGG GCCAAGCCCC	3660
ACTTCCCGGC ACTGTTGTGA GCGCCTCCA GCTCTCTCCA CGCTCTCTGG GTGCCCACAG	3720
GTGCAACGC CAGGCAGGCC CAGCATGCAG TGGCTCTCCC CAAAGCGGCC ATGCTGTGTG	3780
GCTGCCTGCT GCGCCACCCC TGTGGCTCAG GGTCCAGTAT GGGAGCTTCG GGGGTCTCTG	3840
AGGGGCCAGG GATGGTGGGG CCACTGAGAA GTGACTCTGT CAGTAGCGA CCTGGAGTCC	3900
CCAGAGACCT TGTTCAAGAA AGGGAATGAG AACATTCCAG CAATTTTCCC CCCACCTAGC	3960
CCTCCAGGT TCTATTTTGA GAGTTATTTC TGATGGAGTC CCGTGGAGG GAGGAGGCTG	4020
GGCTGAGGGA GGGGGTCTG CAGGGCGGGG GGCTGGGAAG GTGGGGAGAG GCTGCCGAGA	4080
GCCACCCGCT ATCCCCAGCT CTGGGCAGCC CCGGGACAGT CACACACCCT GGCCTCGCGG	4140

32

CCCAGCTGG CAGCCGTCTG CAGCCACAGC TTATGCCAGC CCAGGTCCAG CCAGACACCT 4200  
GAGGGACCCA CTGGTGCCTT GGAGGAAGCA GGAGAGGTCA GATGGCACCA TGAGCTGGGG 4260  
CAGGTGCAGG GACCCTGGCA GCACCTGGCA GGGCCTCAGA ACCCATGCCT TGGGCACCCC 4320  
GGCCATGAGG CCCTGAGGAT TGCAGCCCAA GAGAAGCAGG GAACGCCAGG GCCACAGGGG 4380  
CAGAGACCAG GCCAGGGTCC CTTCGGGCCC TTAGCCCACC CCCTCCAGT AAGCAGGGGC 4440  
TGCTTGGCTA GGCTTCCTTT TGCTACAGAC CTGCTGCTCA CCCAGAGGCC CACGGGCCCT 4500  
AGTGACAAGG TCGTTGTGGC TCCAGGTCCT TGGGGGTCCT GACACAGAGC CTCTTCTGCA 4560  
GCACCCCTGA GGACAGGGTG CTCGCTGGG CACCCAGCCT AGTGGGCAGA CGAGAACCTA 4620  
GGGGCTGCCT GGGCCTACTG TGGCCTGGGA GGTGAGCGGG TGACCCTAGC TACCCTGTGG 4680  
CTGGGCCAGT CTGCCTGCCA CCCAGGCCAA ACCAATCTGC ACCTTTCCTG AGAGCTCCAC 4740  
CCAGGGCTGG GCTGGGGATG GCTGGGCCTG GGGCTGGCAT GGGCTGTGGC TGCAGACCAC 4800  
TGCCAGCTTG GGCCTCGAGG CCAGGAGCTC ACCCTCCAGC TGCCCCGCCT CCAGAGTGGG 4860  
GGCCAGGGCT GGGCAGGCGG GTGGACGGCC GGACACTGGC CCCGGAAGAG GAGGGAGGCG 4920  
GTGGCTGGGA TCGGCAGCAG CCGTCCATGG GAACACCCAG CCGGCCCCAC TGCACGGGT 4980  
AGAGACAGGC GC 4992

## (2) INFORMATION FOR SEQ ID NO:11:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 10 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS:  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: C-terminal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

Xaa Xaa Gly Ser His His His His His  
1 5 10

## (2) INFORMATION FOR SEQ ID NO:12:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 27 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "DNA for bridge peptide"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

AAGAAGAAGC GGCCGCGAAA GAAGAAG

27

## (2) INFORMATION FOR SEQ ID NO:13:

33

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 27 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: other nucleic acid
  - (A) DESCRIPTION: /desc = "DNA for bridge peptide"

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

AAGAAGAAGC GATCGCGAAA GAAGAAG

27